Award Number: W81XWH-€Î ÌÌ̈ÌÏ F

TITLE: Ù˘¦* æ&æ¦Á^&@}[|[*ˆÁ̈̈Qc^*¦æǣ¦}Á̧ã̃ã@ÁV[[|•Á¦¦ÁÔ[*}ã̃ã̧̃^ÁP˘{ æ̧}Á̈Cæ&¦•ÁÇÙVQÝÔPD

PRINCIPAL INVESTIGATOR: Y È̈ÁÓ¦^}ã̧Á^æ¦^•È̈ÁÚ@È̈Ö̈È

CONTRACTING ORGANIZATION: University ofÁS^}č&\ˆ
ÆÚ̈^¢æ̧]*}È̈ÁSŸÁ̈́€Í̈€Î Á

REPORT DATE: U&çà^¦Á̈G€F€

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
              Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for public release; distribution unlimited

REPORT DOCUMENTATION PAGE

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE (DD-MM-YYYY) 01-10-2010	2. REPORT TYPE Annual	3. DATES COVERED (From - To) 25 SEP 2009 - 24 SEP 2010

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Surgical Technology Integration with Tools for Cognitive Human Factors (STITCH)	5b. GRANT NUMBER W81XWH-06-1-0761
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) W. Brent Seales, Ph.D.	5d. PROJECT NUMBER
	5e. TASK NUMBER
E-Mail: seales@uky.edu	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) University of Kentucky Lexington, KY 40506	8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The need for accurate assessment in surgical training has become even more important with the development of new surgical technologies, many of which have transformed methods of treatment for both the patient and the surgeon. Difficult-to-master technologies such as the components of Minimally Invasive Surgery (MIS) highlight the need for surgical competence but do not inherently provide a solution for how to define and measure it. The introduction of new technology such as enhanced visualization, robotics or "plug and play" with closed loop control into the surgeon's art likewise risks increasing danger to the patient or practitioner by introducing complex technology that may confound the goal of simplification or increased safety being sought. The long term goal of this research is to build an integrated surgical technology environment designed for the continuous monitoring of task performance, with a particular focus on the inclusion of important but currently overlooked cognitive measures.

15. SUBJECT TERMS
Subject terms not provided.

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON USAMRMC
a. REPORT U	b. ABSTRACT U	c. THIS PAGE U	UU	19	19b. TELEPHONE NUMBER (include area code)

Introduction

The need for accurate assessment in surgical training has become even more apparent with the development of new surgical technologies, many of which have transformed methods of treatment for both the patient and the surgeon. Difficult-to-master technologies such as the components of Minimally Invasive Surgery (MIS) highlight the need for surgical competence but do not inherently provide a solution for how to define and measure it. The introduction of new technology such as robotics or "plug and play" with closed-loop control into the surgeon's art likewise risks increasing danger to the patient or practitioner by introducing complex technology that may confound the goal of simplification or increased safety being sought. The long-term goal of this research is to build an integrated surgical technology environment designed for the continuous monitoring of task performance, with a particular focus on the inclusion of important but currently overlooked cognitive measures.

Evaluation of surgical skill in MIS can be made more accurate, objective, and general by considering cognitive and environmental factors such as mental workload, stress, situation awareness, and level of comfort with complex tools. To date, our research has shown that a comprehensive framework for measuring cognitive human factors in MIS settings provides an important, statistically significant set of (largely overlooked, in this domain) non-redundant metrics for evaluating performance in the context of new technologies, tasks, and learning methodologies.

Software engineering research efforts will build on the general-purpose Plug-and-Play (PnP) framework and application-specific tools created by previous project effort. Tools and techniques will be developed to take the next step in PnP tools, toward safe and reliable closed loop control. Well-defined methodology will be incorporated in the development process to insure required safety, reliability and robustness attributes for the problem domain. Cognitive studies will build on adapted, validated cognitive metrics in order to address important questions involving diagnosticity, team-level evaluation, and non-intrusive, naturalistic measures of cognitive strain.

The STITCH project has made significant progress in the development of specifications, designs, and implementations of an integrated surgical training and assessment framework and is providing assessment results for specific cognitive measures, including validity and predictive studies. These results are useful for implementing improvements in training methods that seek to use valid cognitive measures as part of the assessment strategy.

Research Accomplishments

1. Cognitive Ergonomics

We made major progress in two human studies this project year – one focused on the refinement of our "time estimation workload" methodology, and the other assessing the relative usability of a global view "dual display" system as compared to a traditional single laparoscopic surgical view. In both studies several rounds of pilot testing were required, including program modification in response. As a result, data collection for these studies was progressive as the pilot tests helped us converge on the correct experimental design. This process is captured in the publications we have prepared to report our progress. We published findings as well as reports on the technical "scaffolding" required to collect the data (fully-automated vocal TLX, for example). The following sections provide more detail about these study designs and our findings.

Time Estimation: Secondary Task Workload Assignment

The goal of the ongoing research on this topic is to create a secondary task measure of workload that is relatively nonintrusive and easily implemented with minimal special equipment. The secondary task involves asking the surgeon or other research participant to make a response (e.g., say "time") every time he or she believes a specified period of time has elapsed. Our research builds upon previous evidence that such "interval productions" become both longer and more variable when a participant is experiencing higher levels of mental workload. However, we are trying to refine the methods to determine the interval, task emphasis instructions, and performance metrics that lead to the greatest workload sensitivity. The goal is to produce an end product that is an automated method of administering the task, recording responses, and calculating a workload score for each experimental trial.

We are completing data collection on our time estimation secondary task metric of mental workload. We have established that time estimation performance trades off with performance of simple surgical transfer training tasks, indicating that time estimation meets the minimal requirements for workload sensitivity. We have also completed partial data collection on two follow-up studies. One of these determines if the sensitivity identified by direct attentional manipulation generalizes to detecting differences attributable to task difficulty. The second study explores whether time estimation is sensitive to more cognitively demanding tasks (and is not restricted to simple visuo-motor surgical tasks).

Comparison of Fully-Automated Vocal TLX vs. Traditional Administration

We have continued to develop a fully-automated vocal TLX. In particular, we were delayed in completing our planned study of attention allocation and interval length effects during the last quarter of 2009 due to limitations in the accuracy of our voice recognition system. The system had an unacceptable false negative rate for some female voices. We subsequently made modifications, conducted a formal evaluation of the system's accuracy, and used the system to produce results during this project year. During the development phase, we collected data in the planned time estimation study using audio-taped backup files to allow manual correction for system misses. Although we have previously shown that non-automated vocal administration of the TLX and MRQ produce acceptably similar results for medical students when they assess workload in basic laparoscopic skills tasks, the fully automated version of the vocal TLX is now in place and allows assessment of the accuracy of the speech recognition capacity when used in a realistic experimental environment. This developed software environment for a fully-automated and formally evaluated system was a major step for improving our experimental setup.

Multimodal Display Evaluation (Assessment)

Our plan to evaluate the "dual display" simulation [the technology is discussed in the tools section] against a traditional single-view laparoscopic display for several navigation tasks was initially delayed due to a combination of problems with our voice recognition system, described above, and refinement of testing scenarios that were the results of several iterations of pilot tests completed during the previous project year. Actual data collection was eventually completed during this project year, using our entire complement of cognitive ergonomics tools. Thus, we have collected subjective and secondary task measures of workload as well as eye tracking and task performance data. As part of this effort, methods of automatically summarizing the eye tracking data have been developed. We will be collecting data for a follow-up study in which the camera navigation task being studied will include a greater planning component (tracking visual landmarks using the shortest routes). The new task is one where the secondary (global) display is more likely to be spontaneously used by participants.

Situation Awareness Assessment

We are developing a method of assessing situation awareness that involves testing for v2p (virtual-to-physical) transfer of training. As part of this process, we specified a testing device in which the 3D surface models rendered in the dual display are also used to generate solid "prints" that can be held and annotated by participants. Their task will be to indicate on the physical models where they found tumors or other targets on the visualized models. A device has been developed to allow the annotated physical model to be measured for congruence with the original model (i.e., to assess the accuracy of the participant's understanding and recollection of the location of critical landmarks). The plan is to implement a new condition in the dual display study in which there is one unannounced trial in which we ask participants to indicate on the physical model what they remember from their most immediate experimental trial.

We have been able to compare the situation models derived from the single laparoscopic view with those derived when the global view is available. This initial data has guided the design, because this design is by necessity a between-group experimental design (i.e., we can only explore participants' memories for target location without warning once during a stimulus sequence).

Remaining work: Secondary Task Workload Assessment Goals

We have completed data collection on all but one experiment in our program of work on time estimation as an appropriate secondary task workload assessment technique for surgical environments. We have clearly established that time estimation performance trades off with performance of simple surgical transfer training tasks, indicating that time estimation meets the minimal requirements for workload sensitivity. A manuscript based on these data has been completed.

We have also completed data collection on two follow-up studies. One of these determined if the sensitivity identified by direct attentional manipulation generalizes to detecting differences attributable to task difficulty. The results suggested that the prior (attentional priority) results do generalize to the detection of differences resulting from a task difficulty manipulation (i.e., differences in the precision required by a peg transfer task). In both cases, a greater investment of resources in the primary task (whether due to attention allocation instructions to the participant or to greater task difficulty) resulted in more variable time estimation performance. The second study explored whether time estimation can also be used to assess workload in surgical tasks with greater cognitive demands (and is not restricted to simple visuomotor "surgical skills" tasks). These data are currently being analyzed.

A final study will test the utility of a pedal-input version of the interval production task as an alternative to the current vocal production procedures. This procedure, if valid, could provide another means of assessing workload in environments where manual control is impossible (i.e., a participant cannot directly respond with a key press because of primary task demands) and where vocal control may be difficult because of environmental factors (background noise level) or lack of appropriate voice recognition software.

Remaining Work: Dual Display Evaluation

During this reporting period, the findings regarding the usefulness of the dual-view surgical display were presented at the Human Factors and Ergonomics Society 54th annual meeting. Preliminary results suggest a benefit for the dual-view display, as compared with a single-scope view, for workload in search-and-traverse tasks and accuracy in creation of 3-D mental models based on the same tasks. This is rather surprising because this task was meant primarily as a baseline task in which we would expect minimal advantages for the dual display; we would mainly hope that it did not impede performance relative to the traditional laparoscopic view. However there was also some evidence of a cost for having the global view available (dual display condition). Performance was poorer (i.e., search time was longer) for trials in which participants had to search for a large number of targets (6 targets per displayed organ). One possibility is that some participants may have relied on the global view when it was not necessary. This finding is consistent with a pattern termed "naïve realism" by human factors investigators, referring to the tendency of some users to prefer displays that look more realistic, for example those that have a three-dimensional look, even when such displays actually compromise performance. Data to understand when and how participants are utilizing the global view on the dual-view display have been collected using eye-tracking procedures and are currently being analyzed.

Data collection is ongoing in a second study evaluating the dual display. In this study, participants perform a task that requires more planning and global spatial comprehension, suggesting a more important potential role for the dual view display. The participants' task is one that requires initial unguided search for targets and then an attempt to traverse the shortest route among those same targets. Thus far, data collection is complete for eight participants, with plans for collecting data from twelve more. A third study is planned to explore which types of information are most important to include in the global view when the information in the scope view is more restricted than is currently the case in this simulation (making it more consistent with actual surgical practice).

We will continue to use the dual-view display simulation for studies through the completion of the project period. The simulation provides a means of testing a variety of issues, including the potential benefits of dual view displays on team performance and on the possible benefits for surgeons of including decisional guidance or just-in-time training on selection of different combinations of display formats to suit their current task demands.

Presentations, Publications, Outreach

During this project year we submitted a number of manuscripts for peer review and subsequent publication. One primary venue for ergonomics studies is the Human Factors and Ergonomics Society's annual meeting, to which we submit research work each year. These and other papers represent the analysis and interpretation of data collected during the last project year and the current project year.

During the project year, we received feedback on the manuscripts we submitted for publication as proceedings papers at the Human Factors and Ergonomics Society's annual meeting. We also received feedback on our revision of the submission to *Applied Ergonomics*. What follows is a list of these papers, annotated to summarize findings and to describe their current status. Copies of each paper are available on the project website as well as through the publication venue.

Also, please note that the Lio et al. paper was conducted as a collaboration, based on research needs identified by the STITCH project, although data collection was not supported by STITCH.

Carswell, C.M., Lio, C.H., Grant, R., Klein, M.I., Clark, D., Seales, W.B, and Strup, S. (in press). Hands-free administration of subjective workload scales: Acceptability in a surgical training environment. Applied Ergonomics, 42, 138-145.

Carswell, C.M. Innovations in Surgical Visualization for Laparoscopic Surgery.

❖ Research presentations at the University of Kentucky's cognitive psychology colloquium series (Oct., 2009) and at the University of Kentucky's instructional systems design colloquium series (Nov., 2009).

Will Seidelman, W., Lio, C., Carswell, C. M., Field, M., Grant, R., Sublette, M., Clarke, D., and Seales, W.B. (2010). Potential Performance Costs associated with Large-Format Tiled Displays for Surgical Visualization. In Proceedings of the Human Factors and Ergonomics Society 54th Annual Meeting (pp. 1430-1434). Santa Monica, CA: Human Factors and Ergonomics Society.

❖ Twenty-five participants performed a surgical training task on a large format display created from one projector or by tiling the images from a 4-, or 9-projector array. Utilizing a large-format display consisting of tiled projector images brings the potential benefits of increased display size with the potential threats to performance of inherent visual artifacts. The effect of these artifacts on performance and subjective workload was assessed. Results indicate that while display size did not affect performance on the surgical task, differences in mental workload were observed. Although a global measure of workload indicated that the tiled displays were the least demanding to use, participants reported deploying additional but highly specific cognitive resources when using these same displays. Their resource shifts seemed to involve adjustments to the perceived control gains created by enhanced size and also the degraded ability to compare target sizes in the larger display, possibly due to the obscuring effect of tile edges.

Sublette, M., Carswell, M., Grant, R., Seidelman, W., Clarke, D., & Seales, B. (2010). Anticipating Workload: Which Facets of Task Difficulty are Easiest to Predict? In Proceedings of the Human Factors and Ergonomics Society 54th Annual Meeting (pp. 1704-1708). Santa Monica, CA: Human Factors and Ergonomics Society.

❖ Prospective workload measures are used to predict how difficult tasks are to complete and how well participants expect to perform. In this study forty-four participants used the NASA-TLX subjective workload assessment to predict the difficulty of surgical training tasks. The goal of the study was to determine the accuracy with which participants could predict task difficulty and whether assessing tasks before performing them affected post-performance judgments. Results showed that participants' prospective judgments were consistent with retrospective judgments of initial performance, except for the underestimation of physical demand. However, after only minimal practice retrospective judgments deviated from both the initial predictions of the experimental group and the initial retrospective assessments of the control group, with mental demand being particularly challenging to anticipate. No significant differences were found between the control and experimental conditions for post-performance assessments suggesting that pre-performance assessment of workload has no effect of post-performance judgment of task difficulty.

Sublette, M., Carswell, M., Han, Q., Grant, R., Lio, C. H., Lee, G., Clarke, C.M., and Seales, W.B. (2010). Dual-View Displays for Minimally Invasive Surgery: Does the Addition of a 3-D Global View Enhance the Construction

of Spatial Mental Models? In Proceedings of the Human Factors and Ergonomics Society 54th Annual Meeting (pp. 1581-1585). Santa Monica, CA: Human Factors and Ergonomics Society.

❖ Technological innovations are at the forefront of advances in minimally invasive surgery. Reduced visual and haptic cues, along with frame-of-reference problems with location and scale can cause surgeons to become disoriented. While most laparoscopic surgeries are performed via the use of a limited, single-scope, two-dimensional (2D) view presented on a monitor in the operating room, there is demand for the availability of three-dimensional (3D), global views. Workload, task-completion time and the ability to recreate spatial mental representations were compared between participants who used the current scope-view display and those who used a dual-view display that included both the scope view and a computationally-generated global view. No statistically reliable improvements were found for the dual-view display over the single-view display for any of our criterion measures, although trends were in the direction of a dual-view advantage for workload in all tasks and accuracy in the reconstruction task despite participants' claims that they did not utilize the global view during the experiment. Future research is needed to better understand both the kinds of information available on global views that can enhance performance during surgical tasks and participants' decisions regarding of when to use different perspective views to support their own performance.

Crant, R., Carswell, C.M., Sublette, M., Lio, C.H., Clarke, D., and Seales, W.B. Standardizing a Verbal Interval Production Secondary Task for Workload Assessment: Maximizing Sensitivity through Interval Choice. In submission for publication in the *Proceedings of the Human Factors and Ergonomics Society 54rd Annual Meeting.*

❖ A study was designed to test the performance trade off, as a function of attention allocation, of a surgical task and a verbal production task. Assuming that attention trades off as a function of task difficulty in multitasking environments, performance tradeoffs may be analogous to the sensitivity of tasks proposed as secondary task workload measures. The goals of this study were to evaluate variables which may influence the sensitivity of a time production task to workload: mode and method of interval production, the target interval, the manner in which productions are summarized, and instructions provided to participants. Participants produced four target intervals (3, 9, 15, and 21 s) while performing a surgical task. Participants were asked to focus attention on each of the two tasks or to devote their attention equally to both. Performance tradeoffs suggest that shorter intervals and use of dispersion estimates as performance scores result in the greatest workload sensitivity.

Lio, C.H., Carswell, C.M., Strup, S.E., Roth, J.S., and Grant, R. (2010). The Operating Room as Classroom: Learning about the Cognitive Challenges Facing Surgical Trainees. In Proceedings of the Human Factors and Ergonomics Society 54th Annual Meeting (pp. 1571-1575). Santa Monica, CA: Human Factors and Ergonomics Society.

❖ Several of the subjective workload measures that were adapted for surgical environments and validated within the STITCH project were applied to actual OR observations by Cindy H. Lio of Innova Design. Although her work was not supported by STITCH, and is outside the scope of work of the current project, we think it is promising to see that the cognitive ergonomic metrics we have been developing can be applied in the target domain and be informative for future design work and training development. We provided Dr. Lio with cognitive task analysis tools for assessing the use of different executive functioning (higher-order cognitive) demands during laparoscopic surgery. Dr. Lio is currently using our checklist in her task analysis that compares experienced and novice surgeons.

❖ Naturalistic observations of two fifth year surgical trainees' laparoscopic performance revealed specific tasks with which they struggled during several seemingly straightforward surgical procedures. Videos of these tasks were further analyzed using retrospective think aloud protocols and NASA-TLX ratings by both trainees and attending surgeons. A common characteristic of the isolated subtasks is that they seemed to stem from trainees' ill-prepared cognitive skills rather than poor technical skill. Specifically, think aloud reports revealed that the trainees focused attention on immediate urgent tasks and failed to strategically plan for action sequences or manipulations as expressed by the surgeons. NASA-TLX results further showed differences in trainees' perception of their effort, performance, and frustration levels when compared to the attending surgeons' perception of their anticipated workload if performing the same tasks.

These preliminary data suggested a gap between the more experienced surgeons' attention allocation strategies and those of the trainees, and it may indicate the need to place more emphasis on cognitive skills training such as multi-tasking during the practice of surgical skills outside the OR.

"Critical Reserves: The Importance of Cognitive Workload in the OR" at the Innovations in the Surgical Environment Conference in Annapolis, MD (March 19).

❖ This was an invited presentation by Dr. Carswell at the annual meeting of the OR of the Future organized by Dr. Adrian Park at the University of Maryland.

Grant, R. (2010). Time Estimation Errors as an Index of Task Demand during Laparoscopic Skills Training: Effects of Target Duration and Attention Allocation. Thesis presented to the Graduate School in partial fulfillment of the M.A. degree. University of Kentucky, Lexington, KY.

❖ The final version of this M.A. thesis was accepted by the University of Kentucky's graduate school in September 2010. The thesis includes a thorough review of the rationale for using time estimation procedures to measure mental workload, as well as results of a study to test the hypothesis that typical laparoscopic skills training tasks and interval production share overlapping cognitive resources (i.e., they show a bidirectional performance tradeoff with changes in emphasis between the two tasks).

Grant, R., Carswell, C.M., Lio, C.H., Clarke, D., and Seales, W.B. (in preparation). Using Verbal Production as an Index of Mental Workload for Assessing Laparoscopic Skills Training.

❖ This is a report of experimental findings aimed at a scientific audience.

Carswell, C.M., Grant, R., Field, M., Lio, C.H., Clarke, D., and Seales, W.B. (in preparation). Time after Time: Using Interval Production to Measure Mental Workload.

❖ This paper is a tutorial in the use of the time estimation procedure directed to practicing human factors engineers in health care settings.

2. *Clinical Assessment at the UMMC MASTRI Center*

The research group at the UMMC MASTRI Center has been collaborating with the University of Kentucky in order to further two studies, both of which require the expertise and laboratory/subject environment established at UMMC. The first study is a validation of the dual display surgical camera navigation system, the technology for which has been developed at UK under this project. In order to facilitate this study, the UK research group has developed the technical system (display, software, hardware) and has delivered an operational system to the UMMC group.

The second study is an ergonomic risk assessment of various positions and techniques using a simulated environment. Again, the UK team supplied technical expertise where possible and necessary, and the UMMC group performed the data collection with experienced surgical subjects. The following sections report on the progress of these studies.

Validation Study Of The Dual Display Surgical Camera Navigation System

During this project year, the role of the research team at the MASTRI center of the University of Maryland (UM) was to receive and test technology centered around the dual display camera navigation system, which has been developed at the visualization center of the University of Kentucky (UK) during this and prior project years. The role of UM is to validate the efficacy of dual display during scope navigation tasks by using a series of cognitive ergonomic assessment tools. The group at UM is positioned to do this evaluation because of the experimental facility established there, the expertise of the professional staff, and the number of available expert subjects who can provide experimental data on the working system.

After establishing a research protocol with the IRB office at the UM, we initiated pilot trials and encountered a few minor problems. A series of phone discussions for the purpose of trouble shooting resolved those issues and led to a second set of trials in order to finalize the experimental setup and study procedures including instructions for subjects. This process also led to a few requests for modification of protocol based on pilot results and reviewer's comments. The modified protocol was approved and the study began recruiting subjects.

The first set of subjects revealed some problems that we feared would jeopardize the study as it had been designed and supported with the technology. First, we discovered that during the task performance in a single display mode, participants pulled the stylus away from the system to increase and improve the camera view. By doing this, participants could get a view similar to the global view in the dual display mode. This didn't model the more realistic situation in surgery - when a laparoscopic camera is pulled away from a target object, the wall of the trocar will *block* the side views of camera as the tip of camera comes into the tube of the trocars. It was clear that this needed to be taken away from the subject in the simulation, to more closely match the reality of surgery. The UK team implemented, in response, a "tunnel effect" into the system, which simulated the real environment.

A second problem was that the global view window should have an improved annotation feature. After the participant identifies the locations of targets and makes an annotation, the global view shows green dots representing the locations. It was still necessary for the sequence of targets, however, to be memorized by participants. This process was leading to an artificially increased cognitive workload. In response to this concern, the UK team implemented a new annotation feature, which leaves the sequence number for each of the targets in addition to the current dot annotation on the global view.

Finally, we identified that, without a proper reminder of using global view during the task performance, participants did not necessarily use the global view. It was crucial to design a system that forced participants to use the global view in order to evaluate its efficacy. This led to a modified experimental protocol in order to remind and encourage participants frequently to use dual display view when it is available.

These protocol changes led to a study with subjects over which we collected data. While we are still finalizing the results, it seems that the global view is efficacious in providing extra information for surgical navigation, without substantially increasing cognitive workload. There was some question in the first round study, however, that the actual tasks were over-simplified. In order to address this, we made improvements to the system:

1) Automatic annotation reminder: once a user correctly marks all the designated target areas by annotations, the system will automatically remind him or her of the next target in the 3D view. Such reminder is not provided in the camera (local) view in order to maintain the fidelity of the camera view to its real-world surgical equivalence. By doing so, we are enhancing the potential functions of a global view by providing useful assisting information to a user without over-simplifying the camera view.

2) The trocar simulation now simulates the tunneling effect caused by moving a surgical trocar and approaches the real complexity of the visual challenge that surgeons have to face in an MIS.

3) Bookkeeping: to allow the system to record the experimental settings, including the environmental parameters so experiments can be duplicated in the future. Log files are also categorized into summary and raw data formats to simplify data analysis tasks.

Dual display publications are in preparation from this data collection at UMMC and UK.

Ergonomic Risk Assessment Of Surgical Techniques And Standing Positions Used By Surgeons Performing Laparoscopic Surgery In A Simulated Environment.

Our previous research data were collected while laparoscopic surgeons performed a laparoscopic cholecystectomy using a virtual reality surgical simulation system from Immersion (software version 1.0). When the simulation software was upgraded to version 2.0, the UMMC research team had concerns as to whether the data, which are already collected using the software version 1.0, and the new data, which will be collected with newer version can be grouped together for further analysis. After investigation, it was found that the newer version significantly changed surgeons' surgical movements due to changes made in tissue-instrument interaction, difficulty task requirement, and the level of visualization. Therefore, it was concluded that further data collection using the newer software would not be beneficial for the project.

Data collection proceeded under this project year and yielded the primary finding that laparoscopic procedures (specifically repetitive cholecystectomy) pose substantial ergonomic risk of physical injury to surgeons. This result is based on data collected from this study and has been presented in publication and at the 2010 SAGES annual conference.

Submitted Manuscripts.

- Youssef Y, Lee G, Godinez C, Sutton E, Seagull FJ, Park A (2010) Laparoscopic Cholecystectomy Poses Physical Injury Risk To Surgeons: Analysis Of Hand Technique And Standing Position, 2010 SAGES Annual Conference.

- A manuscript regarding this research study has been submitted to Journal of Surgical Endoscopy and is in the process of revision.

3.   Tools and Technology

Investigation of the deformation cloning method

During this project year we investigated an extension of the "deformation cloning method." The extension combines multiple deformation models in order to incorporate deformation models dynamically, overcoming the circumstance of limited training data and adapting to more training data whenever available. Unfortunately the departure of Dr. Han, the primary architect of this method, stalled this effort. Until we can replace his expertise we have suspended work in this direction.

The Dual Display

The cognitive ergonomic studies using the custom-developed "Dual Display", which is both a hardware and software system with haptic feedback spanned this project year. Subjective workload metrics based on the NASA TLX and objective measures using the faceLAB eye tracking system analyzed user performance. Staff provided technical support and calibration for the testing system, as well as customized software tools for further study. These tools involved speech recognition, data aggregation from various instruments, and 3D reconstruction of a phantom model.

The dual display software underwent several revisions based on feedback from our human factors team at UK and from our collaborators at the University of Maryland. New interfaces were added for easier management of user studies. The scope view was made more realistic by including the "tunnel vision" of a simulated trocar, to restrict the range of motion of the camera. The code base is currently undergoing revision and cleanup to be handed off to new staff members for further development.

User studies of the Dual Display software continued throughout this project year, and several new features were added to the software at the request of our collaborators at the University of Maryland. Data analysis tools were created for processing and unifying the output of the numerous log files from each component of these studies. The faceLAB software has been updated to version 5.0 to improve reliability.

The VocalTLX speech recognition tool was released for download on the project web site. VocalTLX automates the recording of subject responses to workload assessment questionnaires like the NASA TLX, SSSQ, and MRQ, reducing both labor and human error in the transcription of this data. A companion tool assists in measuring secondary task performance based on time estimation, a technique pioneered by our research team. Both software packages are written in Java and based on the Sphinx4 library from Carnegie-Mellon University. Documentation and download is available at http://halsted.vis.uky.edu/~dan/VocalTLX/.

In summary, we extended the dual-display visualization system during this project year with the following functions and delivered the system to our partner at UMD-Baltimore. The system was used at UMMC to conduct user studies to test the effectiveness of our visualization scheme.

1)   Using a haptic device as a full 3D navigational control on the camera. Our purpose is to allow subjects to browse the 3D objects freely like they would to an MIS scene.

2) Automatically recording the elapsed time for all key events to improve the accuracy of collected experimental data and to relieve the burden on experiment supervisor. The system records the subject's progress and report back whenever a trial (task) is finished.

3) Automatically recording the performance of each subject by logging all the successes and failures in finding specific targets.

4) Adding automatic target recognition to provide necessary feedback for experiment subjects. Each interaction between the subject and the system is processed and appropriate audio cues are provided to the subject.

5) Adding the capability to manually orient the global view and to automatically snatch back to the original view once the global view manipulation is finished.

6) Eliminating any artifacts in our recorded data by letting each subject go through a randomized list of experiments. By this means, we can minimize any biased influence on the subjects' performance from a fixed order of tasks.

7) The system will also be used to test the mental reconstruction ability of subjects with or without the help of the 3D view. After a series of randomly ordered experimental tasks, each subject will be asked to identify the exact positions of the targets by putting artificial markers on a real-world 3D object, directly produced from the 3D model used in the experiments. A computer-vision system will then be used to test the accuracy and reliability of the positions of those markers.

The image below shows a user view from the system. The technical team made a number of improvements to the system to support the collection of subject data at UMMC. In particular, the technical team achieved the following:

1) Simulating the trocar effect: when a user moves away from the main target object far enough, the system will simulate the tunneling effect caused by using a trocar in the real surgery, shown in the figure below. By doing this, we allow the system to simulate the complexity of the visual challenge that surgeons have to face in a real world MIS so we can have much more meaningful results comparing the simulated surgical view with the global view.
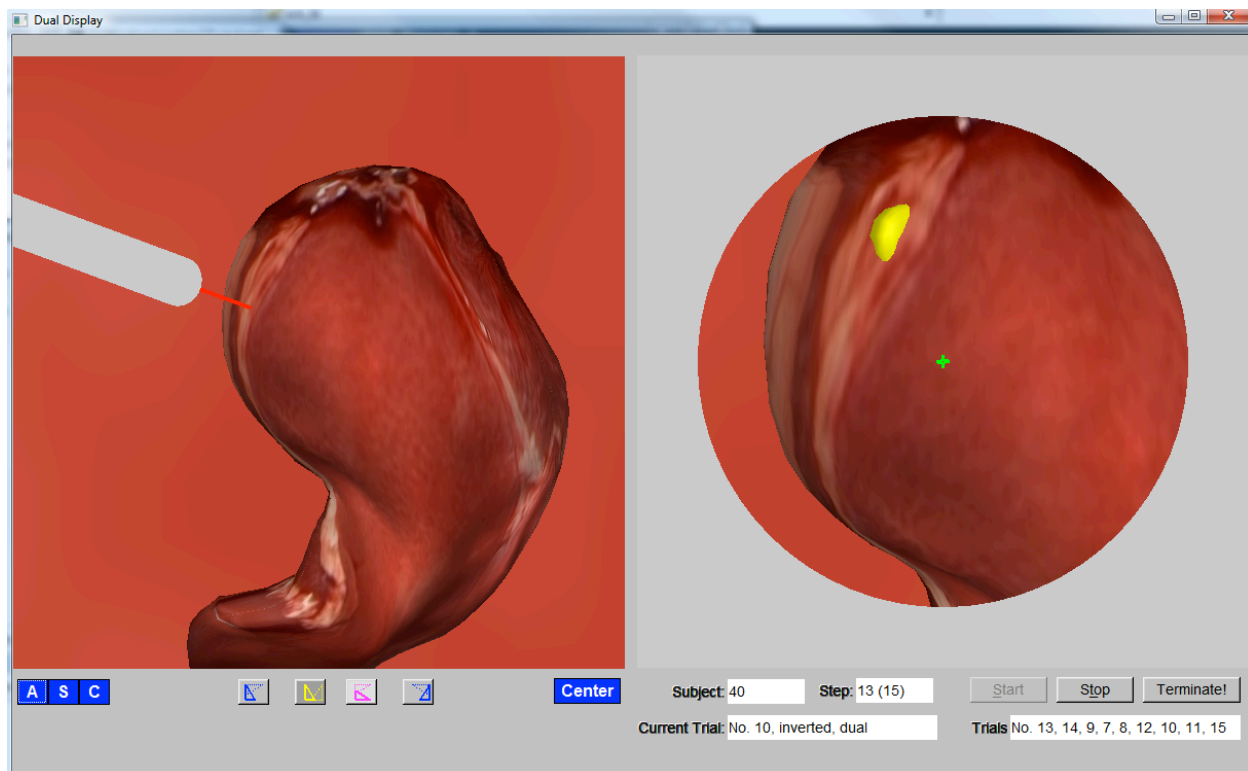
2) Changing the annotation: the prior annotation implementation gave too much information to users when they work in the simulated camera view for MIS, which, in our hypothesis, weakened the purpose of having a global view to assist their performance. By approaching the annotations to the same realistic level as the real world MIS, we hope to have a more accurate evaluation of our system works.

3) Allowing a different ordering and requirements of tasks performed by subjects: by doing so, we are allowing users to take more advantage of the augmentation provided by the global view. Again, we are in hope to acquire a more realistic evaluation of our system.

4) Bookkeeping: to allow the system to record the experiment settings, including the environmental parameters.

Tools Support

Progress has continued during this project year on the adaptation of a software prototype of a stereoscopic measurement tool for broader use. We group this work under "tool support" although enough progress has been made that we report separately in a later section. Several new techniques have been studied this year for automated instrument tracking within the video. New phantom data has been collected using the stereoscope in our lab, and more testing is planned in the coming months using the much higher quality endoscope of the daVinci robot. More on this progress from the year is reported under "Stereo Endoscope Analysis." The PhD student Sami Taha is continuing experiments in this area as part of a PhD thesis that will explore the value of measurement data in monitoring and assessing *automatically* the performance skill of experts who perform tasks.

Projector-Based Display System

The multi-projector display system underwent an upgrade to new hardware during this project year, with a few new added capabilities as well. Full HD video capture is now supported, with only a small increase in latency over standard definition formats. The report on this system (its construction and performance) was published last project year and the software to drive the system is publically available on the project website and by request.

Multi-Modal Imaging

Evaluation of multispectral imaging began with the setup and installation of our multispectral camera system. The camera provides high resolution images across 13 wavelengths of light. Unfortunately no user-ready system exists for laparoscopic applications, so the experimental system has required adaptation for surgical testing. The biggest of these obstacles, beyond the physical size of the camera, are the temporal artifacts introduced by combining multiple photographs taken over a period of several seconds. We have investigated ways of shortening the temporal interval for image acquisition.

In order to evaluate the adapted camera system's performance in enhancing surgical video, we photographed a live mouse undergoing experimentation involving open surgery at the UK Department of Toxicology. This initial survey provided data by which we can guide the adaptation in understanding the capabilities and enhancements possible through imaging in light beyond the visible spectrum.

Further analysis on the multi-spectral images from the mouse surgery revealed slightly more visible vasculature in the blue and ultra-violet bands, though not to the extent and the depth that we had hoped. It is likely that continued research will depend on the use of fluorescent contrast agents to better target specific structures in the anatomy. This initial work was presented in April at the poster session of the World Congress of Endoscopy Surgery, hosted by the Society of American Gastrointestinal and Endoscopic Surgeons.

Safe, Reliable Engineering of PnP Systems

During the project year we worked with researchers from the University of Pennsylvania and Korea University on developing tooling to support the emerging MD PnP systems engineering framework. Previous prototypes had demonstrated that existing tooling was inadequate to the task of analyzing formal models of MD PnP systems architectures. The problems were twofold: (1) Immaturity of the OSATE tool for modeling and analysis of the Architecture Analysis and Design Language, and (2) system architecture constraints on model sizes in the VERSA tool.

Version 2 of the OSATE tool is under development within the Software Engineering Institute at Carnegie Mellon University. A preliminary alpha-release version is currently available, and more stable versions are expected to be released throughout 2010.

A revised version of the VERSA tool has been developed jointly during this project year by researchers from Fremont Associates (a STITCH sub-contractor) and researchers from Penn and KU, funded in part by the STITCH project. The new VERSA tool will eliminate current constraints on input specifications that arise from static flattening of value-passing operations. The language supported will be true ACSR-VP (Algebra of Communicating Shared Resources with Value Passing) with symbolic communication channels. The system is also being re-architected to support distributed processing for analysis tasks, and run-time configurable analysis plug-ins.

Figure 1 illustrates the current working software architecture for the VERSA replacement, to be known as Gabbro. Users will interact with a GUI implemented in Java. Java was selected as the implementation language for the front-end due to portability and look and feel considerations. The back-end analysis will be carried out in a so-called native language, *i.e.,* a language that can be compiled directly into machine code, for the sake of efficiency. An API for plugging back-end analysis tools into the system GUI is being developed. Standard interfaces based on RMI (Java's implementation of remote procedure calls) and JNI (Java's byte-code to native code bridge) are used to facilitate distribution of computing load and an efficient connection between Java and native code.
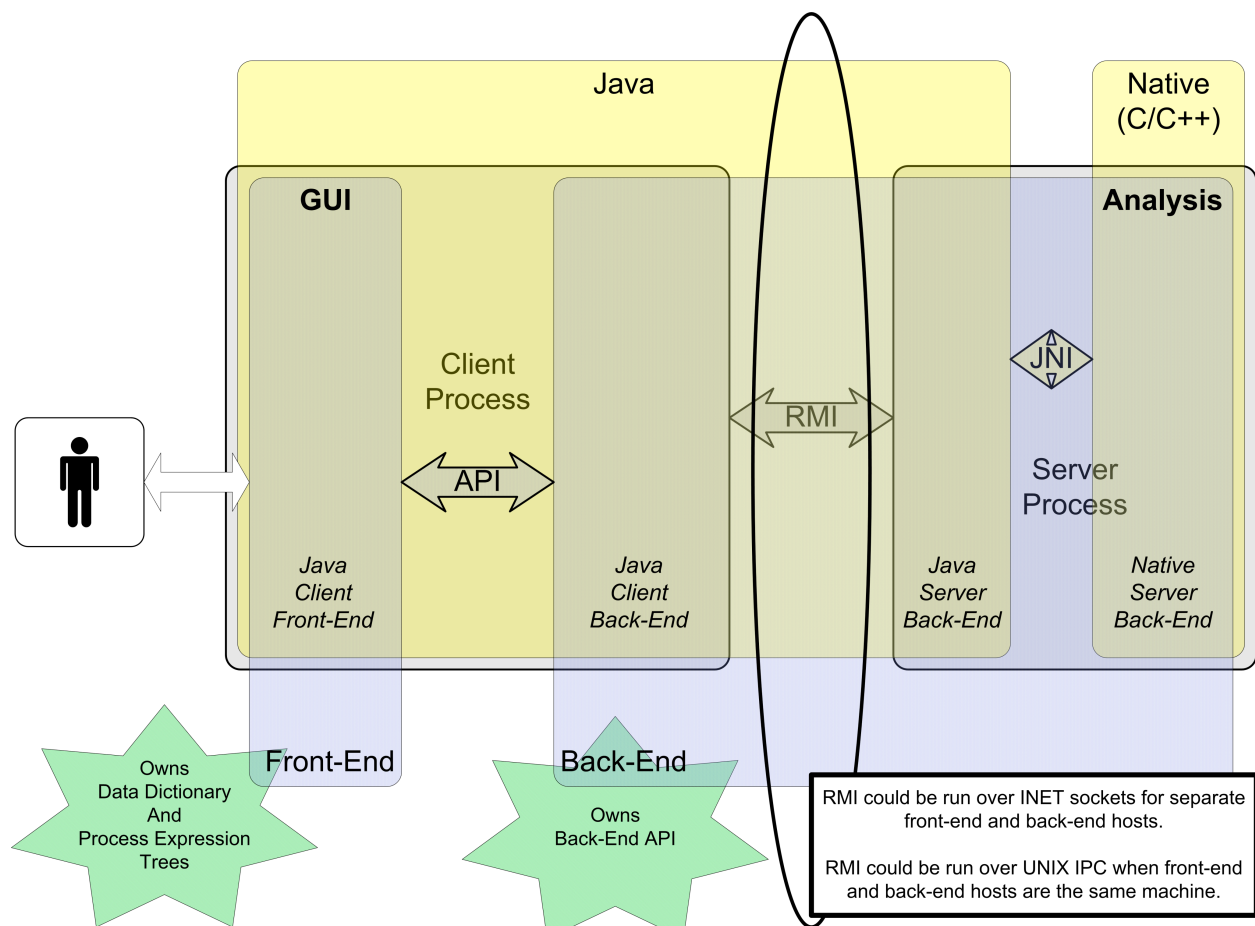

Figure 1: Gabbro System Architecture

The following figure illustrates the current state of the development. A workspace explorer window (upper left pane) is used to manage folders and files containing input specifications. An editor (upper right pane) allows creation and modification of specifications. The current implementation is a straightforward text editor, but the implementation has been carried out in a way that will support advanced semantically rich editing of input specifications. Output of various analysis steps in the processing of specification is shown in the lower pane.
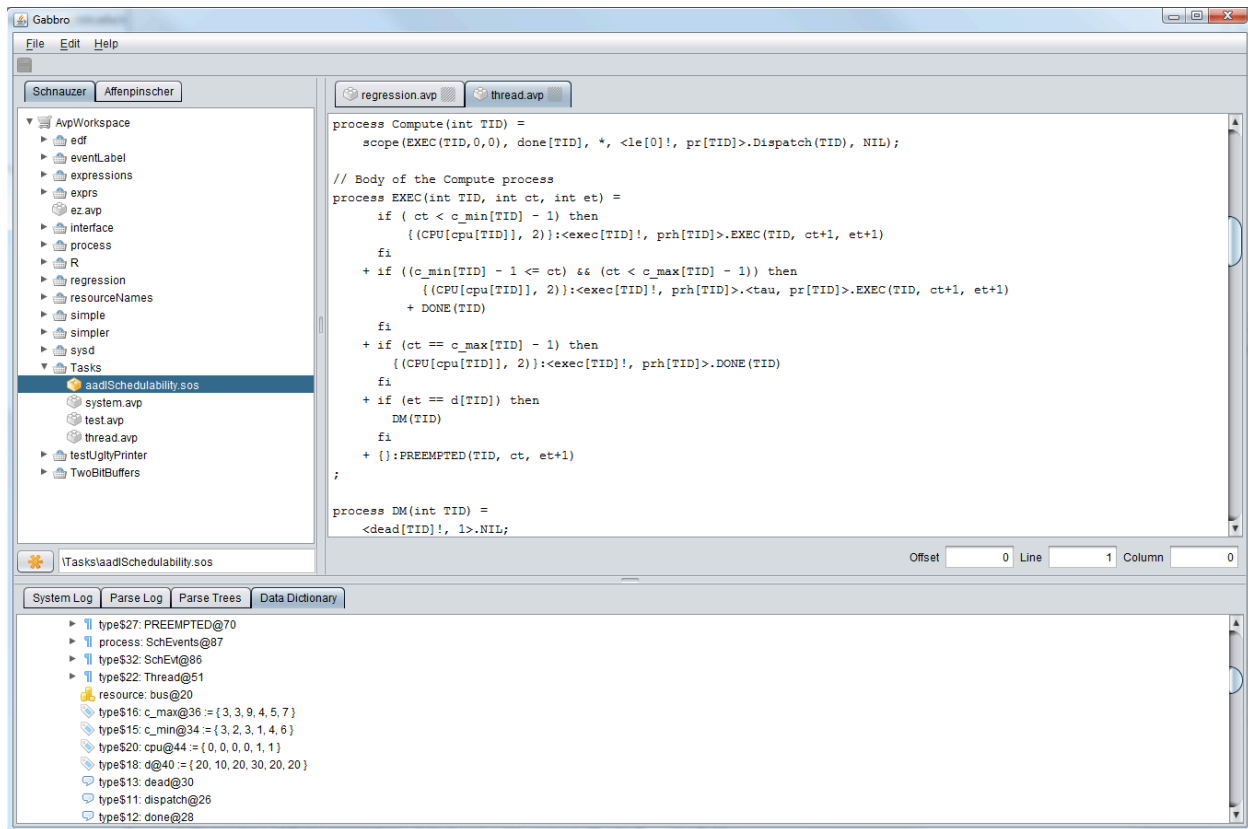
At present the tool parses input specification into an internal representation for use by the tool and plug-ins. The GUI is complete enough to allow the editing and manipulation of specification files, and manual analysis of translation results. The GUI will also allow the selection of compiled elements that are passed as parameters to independently developed plug-ins. Initial plug-in interface testing is complete and independent plug-in development is ongoing.

The long-term goal for the Gabbro tool is to provide an environment that can be used to formally analyze realistic specifications of encoded MD-PnP systems and carry out safety and liveness analysis. The result of that analysis will be fed back to higher-level MD-PnP tools to iteratively develop safe and reliable MD-PnP Systems

Stereo Endoscope Analysis

We tested two varieties of stereo endoscopes with different optical properties. Both endoscopes pack two cameras into a single 10mm endoscope; however, these systems differ greatly in their construction.

1. Single-channel Stereo Scope: Vista Medical Technologies' stereo scope uses a standard endoscopic lens with two CCDs positioned slightly apart sharing the same optical path. The cameras output analog NTSC video signals, with a resolution of 640x480. The video is then routed to a head-mounted display to provide stereoscopic viewing. The disparity between Fig. 1. Stereoscopic endoscopes use two cameras to capture a three dimensional view to minimally invasive surgery the two cameras is small (< 5mm). Since the cameras share a single lens system, each is positioned slightly off-center in relation to the attached endoscope.

2. Bi-channel Stereo Scope: Intuitive Surgical's da Vinci® Surgical System is a robotic surgical platform. As such, it is capable of handling larger and more complex instruments that a human operator might find

unwieldy. It uses a larger camera, manufactured by Olympus, attached to an endoscope with two separate lenses embedded inside a single tube. The da Vinci® Surgical System incorporates high-definition technology at a resolution of 1280x1024 and outputs directly to two precisely aligned LCD monitors, positioned so that each is viewable by only one eye. Due to the complexity of capturing synchronous uncompressed high-definition video, we captured images from its standard definition NTSC video outputs. This also allows us to directly compare the optics of the two systems, as it removes the variable of differing resolutions.

In computer vision, a camera is generally modeled as an ideal pinhole camera. This model describes the mapping of three-dimensional world coordinates to two-dimensional image points via a matrix containing the intrinsic properties of the camera, such as focal length and center of projection. Calibration is the procedure by which we approximate these unknown properties using some known constraints, in this case, images of a calibration target.

To calibrate the intrinsic parameters of each endoscope, we captured 20 images of a checker board moving within the view of a stationary camera. The locations of each corner in the target were isolated, and approximated to sub-pixel accuracy by intersecting the gradient responses of the black and white edges of the pattern. These points, on a target of known measurements, were fed into an optimization routine to derive the camera properties. This procedure was performed several times for each endoscope using different image sets to verify the accuracy of our calibration.

A feature point in the image of one camera can be plotted as a ray in three-dimensional space, originating at the camera center and passing through the image plane to its 3D location. The plane created by this line and the baseline intersecting the two camera centers is called the epipolar plane. The corresponding point in the second image will fall on the epipolar line where the epipolar plane intersects the image plane. This epipolar geometry forms the basis of stereo reconstruction.

In an ideal situation, the two rays back-projected through the image points would intersect at the true three-dimensional location of the feature. Since real images are rarely ideal due to image noise and errors in feature matching, these rays will generally not intersect exactly, but we can approximate the location of the 3D point as the midpoint of the shortest line segment joining the two rays. This method of stereo triangulation is used to approximate the three-dimensional location of feature points.

We again use the same calibration target of known dimensions, to more easily calculate the errors between ground truth and reconstructed data. A second set of images, not used for calibration, is captured for this analysis.

Twenty additional image pairs that were not used for calibration were selected from the output of each endoscope. Each image contained our target checkerboard pattern positioned roughly parallel to the image plane, orthogonal to the Z axis (depth), at increasing distances from the endoscope. Reconstruction by stereo triangulation was performed on each pair of matching feature points. Graphs of the reconstructed patterns are presented below, alongside the original points as calculated from the images.

 The methods developed in this study could be directly applied to create a virtual measuring tape for use during MIS procedures. With further refinements to our software, an endoscope could be calibrated before surgery from video of a calibration target in only a few minutes. High-contrast features in the video, such as the surgical instruments, could be tracked reasonably accurately. By positioning the instruments near anatomy of interest, a surgeon could quickly make real-time measurements of a patient's anatomy. Approximate accuracy of the calculations could also be determined, and could be further improved by restricting measurements to a plane.

The ability to accurately judge anatomical scale has many practical implications to minimally invasive surgery. This is one of the key problems that makes MIS more difficult than open surgery.

Measurement Tool

We conducted another round of data collection using the daVinci Surgical System at the University of Kentucky Hospital in May. In this study, measurements were taken on phantom organ models to simulate the anatomy for a more realistic use case. Significant progress has been made on overall measurement accuracy, and work continues toward fully automated instrument tracking. Another study is planned, with HD video capture to improve accuracy.

Part of the STICH project is to develop a virtual ruler to calculate distance between two points from the stereo video of the laparoscopic cameras. New software is developed to achieve the ruler. The software will calculate the real world distance between the two tips of the surgery tools. It includes several parts. First part is camera stereo

calibration which uses stereo calibration from OpenCV library. The second part is un-distorting and rectifying the left/right images. Third is using the resulting model from the previous step to transform the chosen points from the left/right images into real world coordinates. The last step is to calculate the distance between the two world points. The following is preliminary data of using the software on the Da vinci and Vista laparoscopes. The models used are human phantom models. In the table the test is done via two approaches. The first method is to track the tool tips on several frames and calculate the distance for each frame and average the result. The second method is to calculate the distance between the chosen points and the distances between the surrounding pixels of the chosen points, averaging the results.

Da vinci

Trial	Type	Ground Truth	Mean +/- STD Dev
1	Liver	106mm	93.3 +/- 9.8
1(surrounding pixels)	Liver	106mm	100.6 +/- 0.23
2	Liver	127mm	127.0 +/- 4.4
2(surrounding pixels)	Liver	127mm	126.6 +/- 0.13
3	Liver	99mm	95.5 +/-6.8
3(surrounding pixels)	Liver	99mm	93 +/- 0.1
4	Kidney	57.15mm	52.4 +/- 9.5
4(surrounding pixels)	Kidney	57.15mm	59 +/- 0.1

Vista

1	Kidney	57.15mm	56.257 +/- 13.0
2(surrounding pixels)	Liver	99mm	75.3 +/- 0.8
3(Surrounding pixels)	Lung	114mm	112.1 +/- 0.5
4(surrounding pixels)	Lung	86mm	81.5 +/- 0.2

Error analysis of the 3D data collected thus far has revealed wide variance in the reliability of measurements. Another round of data collection on the daVinci system is planned in order to more rigorously test accuracy in a tightly controlled environment.

Object recognition algorithms are also being studied to automatically locate surgical instruments in the scene. This will enable automatic tracking that will adapt the virtual measurement tool to something that will be practical in a real-world scenario.

4.   Supported Personnel

The following is a list of personnel supported by the research effort in this project year.

Catherine M Carswell	Russell C. Grant
James D Hoskins	Daniel L Staley
William B Seales	Qiong Han
Stephen Strup	Stephen P Bailey
Matthew Douglas Field	Michelle Sublette
Geoffrey James	James Conway
Chris Campbell	Julia Martinez

Kyle Kolpek   Ryan Taylor
Jose Farrington   Elizabeth Lutin
Patricia Calvert   Mikhail Slavin
Jacquelyn Thompson   Sami Taha
Will T. Seidelman   Juan Carlos Valentin
Juan De Dios Rivera

Key Research Accomplishments (Summary)

During this project year we modified cognitive measures and technology to support them in order to collect subject data from two studies at UK, one on mental workload and situation awareness, and the other on secondary task assessment and time estimation. In the process, we created a vocal TLX measure that we have shown to be valid. In partnership with UMMC we have deployed a "dual display" system, which has been improved and used at UMMC for the collection of subject data. The UMMC group has also used our measures to collect subject data on ergonomic factors for surgeons performing repetitive cholecystectomies. Finally, the technical group at the University of Kentucky has made progress in the technology to support these efforts as well as the development of a customized multi-spectral imaging system for a proposed tissue study, the development of PnP software verification tools in the form of the Gabbro (VERSA follow-on) tool, and the study of camera systems as measurement devices to aid future work in performance assessment of surgical tasks.

Each of these accomplishments was reported in quarterly reports and has been summarized in professional publications, reported below. In addition, software systems have been made available via the project website and by request. These systems are configurable and implement (1) the display controller for scalable, HD resolution display systems; (2) the vocal NASA TLX; and (3) the dual display system.

In the final project completion year we anticipate conclusion studies, release of remaining software improvements, and a findings report consisting of major findings summaries and media summaries for a more general audience with pointers to specific data and professional papers to back up findings and claims.

Reportable Outcomes

ABSTRACT ACCEPTED

- Youssef Y, Lee G, Godinez C, Sutton E, Seagull FJ, Park A (2010) Laparoscopic Cholecystectomy Poses Physical Injury Risk To Surgeons: Analysis Of Hand Technique And Standing Position, 2010 SAGES Annual Conference.

FULL PAPERS ACCEPTED

- Carswell, C.M., Lio, C.H., Grant, R., Klein, M.I., Clark, D., Seales, W.B, and Strup, S. (in press). Hands-free administration of subjective workload scales: Acceptability in a surgical training environment. Applied Ergonomics, 42, 138-145.

- Will Seidelman, W., Lio, C., Carswell, C. M., Field, M., Grant, R., Sublette, M., Clarke, D., and Seales, W.B. (2010). Potential Performance Costs associated with Large-Format Tiled Displays for Surgical Visualization. In Proceedings of the Human Factors and Ergonomics Society 54th Annual Meeting (pp. 1430-1434). Santa Monica, CA: Human Factors and Ergonomics Society.

- Sublette, M., Carswell, M., Grant, R., Seidelman, W., Clarke, D., & Seales, B. (2010). Anticipating Workload: Which Facets of Task Difficulty are Easiest to Predict? In Proceedings of the Human Factors and Ergonomics Society 54th Annual Meeting (pp. 1704-1708). Santa Monica, CA: Human Factors and Ergonomics Society.

- Sublette, M., Carswell, M., Han, Q., Grant, R., Lio, C. H., Lee, G., Clarke, C.M., and Seales, W.B. (2010). Dual-View Displays for Minimally Invasive Surgery: Does the Addition of a 3-D Global View Enhance

the Construction of Spatial Mental Models? In Proceedings of the Human Factors and Ergonomics Society 54th Annual Meeting (pp. 1581-1585). Santa Monica, CA: Human Factors and Ergonomics Society.

- Crant, R., Carswell, C.M., Sublette, M., Lio, C.H., Clarke, D., and Seales, W.B. Standardizing a Verbal Interval Production Secondary Task for Workload Assessment: Maximizing Sensitivity through Interval Choice. In submission for publication in the *Proceedings of the Human Factors and Ergonomics Society 54rd Annual Meeting.*

- Lio, C.H., Carswell, C.M., Strup, S.E., Roth, J.S., and Grant, R. (2010). The Operating Room as Classroom: Learning about the Cognitive Challenges Facing Surgical Trainees. In Proceedings of the Human Factors and Ergonomics Society 54th Annual Meeting (pp. 1571-1575). Santa Monica, CA: Human Factors and Ergonomics Society.

- Grant, R. (2010). Time Estimation Errors as an Index of Task Demand during Laparoscopic Skills Training: Effects of Target Duration and Attention Allocation. Thesis presented to the Graduate School in partial fulfillment of the M.A. degree. University of Kentucky, Lexington, KY.

MANUSCRIPTS IN PREPARATION

- Q. Han, S. M. Strup, M. Carswell, D. Clarke, and W. B. Seales, Verification of a Model Completion Method for Minimally Invasive Surgery (MIS).

- Grant, R., Carswell, C.M., Lio, C.H., Clarke, D., and Seales, W.B. (in preparation). Using Verbal Production as an Index of Mental Workload for Assessing Laparoscopic Skills Training.

- Carswell, C.M., Grant, R., Field, M., Lio, C.H., Clarke, D., and Seales, W.B. (in preparation). Time after Time: Using Interval Production to Measure Mental Workload.

- Q. Han, S. M. Strup, M. Carswell, D. Clark, and W. B. Seales, A Statistical Deformation Model of Multiple Subjects for Minimally Invasive Surgery (MIS), to be submitted to MICCAI 2010.

- Q. Han, et al., A Generalized Medial Shape Representation, to be submitted to MICCAI 2010.

- Q. Han, D. Clark, W. B. Seales, Use of Inertia Measurement Unit (IMU) in Computer-Aided Minimally Invasive Surgery (MIS), to be submitted to ISVC 2010.

- Q. Han, S. M. Strup, M. Carswell, D. Clark, W. B. Seales, An integrated Visualization System for Computer-Aided Minimally Invasive Surgery (MIS), to be submitted to ISVC 2010.

Conclusion

During this project year the STITCH project has made significant progress in all of its objective areas. (1) Assessment of task cognitive ergonomics by psychology researchers has yielded significant results, many published, and many yet to be published, in demonstrating the validity of our methods. These methods have also been applied to realistic studies using medical students and actual clinical equipment. (2) Significant progress was made in the deployment and analysis of tools and techniques in the MASTRI center at the University of Maryland Medical Center. Studies were carried out using medical students and residents in the MASTRI center, yielding significant, published results. In particular, the "Dual Display" environment was deployed, improved, and used to collect subject data (3) The multi-modal imaging system was acquired and adapted for an initial study to reveal vasculature if possible. Other technical tools were also developed, either for direct application in a test-bed setting, or to support the other efforts.   And (4) systems architecture for safe, reliable engineering of PnP OR systems has been created and is guiding ongoing tool building efforts.